

Naar een nieuwe fase in het longitudinale onderzoek

Citation for published version (APA):

Allen, J. P., Heijke, J. A. M., & van der Velden, R. K. W. (1999). *Naar een nieuwe fase in het longitudinale onderzoek*. Researchcentrum voor Onderwijs en Arbeidsmarkt, Faculteit der Economische Wetenschappen. ROA Reports No. 4 <https://doi.org/10.26481/umarep.1999004>

Document status and date:

Published: 01/01/1999

DOI:

[10.26481/umarep.1999004](https://doi.org/10.26481/umarep.1999004)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Naar een nieuwe fase in het longitudinale onderzoek

**Preadvies, samengesteld op verzoek van de
Programmaraad Onderwijsonderzoek van NWO**

ROA-R-1999/4

Jim Allen
Hans Heijke
Rolf van der Velden

Researchcentrum voor Onderwijs en Arbeidsmarkt

Faculteit der Economische Wetenschappen en Bedrijfskunde
Universiteit Maastricht

Maastricht, april 1999

ISBN 90-5321-251-5
Sec98.208/RvdV

Inhoud

	Bladzijde
Samenvatting	i
1 Inleiding	1
2 Huidige en vroegere cohorten	4
3 'Opbouw, onderhoud en slijtage van menselijk kapitaal'	8
4 De (on)mogelijkheden van cohortonderzoek in vergelijking tot andere databronnen	11
5 De data-infrastructuur in de toekomst	15
6 Commentaar van experts op het preadvies: hoofdpunten	21
Bijlage 1: Lijst van gevoerde gesprekken	24
Bijlage 2: Geraadpleegde documentatie	25
Bijlage 3: Commentaar van experts op het preadvies	26

Samenvatting

In het advies wordt een ontwerp geschetst hoe op de middellange termijn in Nederland de infrastructuur van het longitudinale onderwijsonderzoek zou kunnen worden ingericht. Een belangrijk uitgangspunt daarbij is dat de huidige onderwijscohorten deel uitmaken van een breder onderzoeksterrein, dat wordt aangeduid als 'het proces van opbouw, onderhoud en slijtage van menselijk kapitaal'. Dit proces beperkt zich niet tot het onderwijs, maar vindt plaats in verschillende fasen over het hele leven van mensen heen. Elke levensfase kenmerkt zich door verschillende accenten met betrekking tot de doorgemaakte ontwikkeling, de gemaakte keuzes en de relevante contexten.

Verschiedende vormen van onderzoek worden besproken in termen van hun bruikbaarheid voor het onderzoek naar de opbouw en onderhoud van menselijk kapitaal. De bruikbaarheid van cross-sectioneel onderzoek en registraties blijkt tamelijk beperkt in dit verband. Levenslange cohorten zijn niet erg geschikt voor onderzoek dat zich beperkt tot een bepaalde onderwijsfase. Fase-specifieke cohorten daarentegen missen het lange termijn perspectief. Bij beide vormen van cohortonderzoek is bovendien de representativiteit van de steekproef ten opzichte van de landelijke populatie moeilijk te garanderen. Om deze redenen wordt bepleit om op de middellange termijn te komen tot een nieuwe infrastructuur waarbij verschillende databronnen worden verankerd in een integrale landelijke basisregistratie. Als basis hiervoor wordt het waarschijnlijk binnenkort in te voeren onderwijsnummer voorgesteld. Deze basisregistratie zou achtergrondgegevens en prestatie-indicatoren van leerlingen moeten bevatten. De basisregistratie zou bovendien als landelijk steekproefkader kunnen dienen voor cohortonderzoeken en schoolonderzoeken. Deze vormen een onlosmakelijk complement op de basisregistratie vanwege de noodzakelijk geachte verdieping van gegevens. De integratie van de verschillende databronnen op basis van een landelijke basisregistratie maakt dat niet zozeer sprake is van een 'onderzoek', maar dat eerder sprake is van een 'data-infrastructuur', waarbij voor verschillende onderzoeksvragen verschillende combinaties van gegevens gebruikt kunnen worden. De infrastructuur zou niet alleen moeten worden gebruikt en onderhouden door instanties die zich bezighouden met onderwijsonderzoek, maar juist door een verscheidenheid aan belanghebbende organisaties. De voorgestelde infrastructuur lijkt daarmee niet alleen een aantal zwaktes van de huidige afzonderlijke onderzoeken te kunnen overwinnen, maar maakt het ook mogelijk om de efficiëntie te verbeteren en meer aandacht te schenken aan de sterke kanten van de verschillende vormen van dataverzameling.

We hebben het niet tot onze taak gerekend hier een complete blauwdruk te leveren voor een ideale organisatorische opzet. Het heeft pas zin om de organisatorische en financiële details verder uit te werken als het duidelijk is dat er voldoende draagvlak is voor het idee. Wanneer dit het geval is, zal ervoor moeten worden gezorgd dat de

koppeling van verschillende informatiebestanden aan het onderwijsnummer mogelijk wordt gemaakt, en dat de data-infrastructuur toegankelijk wordt – via een vertrouwelijke sleutel – voor gebruik voor legitieme onderzoeksdoelen. Vervolgens zal een stappenplan moeten worden ontwikkeld en elke stap nader worden uitgewerkt.

De finale versie van het preadvies is in november 1998 gestuurd aan zes experts, met het verzoek om commentaar. In paragraaf 6 zijn de belangrijkste punten die door de experts naar voren zijn gebracht kort samengevat. Alle experts onderschrijven het idee van een data-infrastructuur waarbij gegevens uit verschillende bronnen aan elkaar worden gekoppeld via een volledige registratie. Ze plaatsen ook een aantal kanttekeningen bij het door ons voorgesteld ontwerp. Het commentaar concentreert zich vooral op vier belangrijke issues: privacy, toegankelijkheid, het gebruik van het onderwijsnummer en de organisatorische opzet van de data-infrastructuur.

1 Inleiding

Dit preadvies is samengesteld op verzoek van de Programmaraad Onderwijs-onderzoek (PROO) van NWO. In het preadvies wordt een ontwerp geschetst hoe de data-infrastructuur rondom het onderwijs-cohortonderzoek in Nederland in de toekomst zou kunnen worden ingericht. Directe aanleiding hiervoor wordt gevormd door het feit dat de PROO de functie van programmering van het onderwijsonderzoek van SVO heeft overgenomen. Binnen deze onderzoeksprogrammering nemen de cohort-onderzoeken een belangrijke plaats in. Met de overgang van SVO naar NWO komt de potentiële wetenschappelijke betekenis van cohortgegevens sterker dan ooit naar voren. Een tweede aanleiding is de geplande invoering van het persoons-gebonden onderwijsnummer, dat op termijn bijzondere consequenties kan hebben voor de opzet en inrichting van het cohortonderzoek.

Als voorbereiding op deze taak hebben we een verkenning verricht op basis van een documentenanalyse en gesprekken met direct betrokkenen en andere experts. De documenten hadden betrekking op cohortonderzoek in binnen- en buitenland, en op een aantal nieuwe ontwikkelingen zoals de invoering van het onderwijsnummer. We hebben gesproken met onderzoekers die betrokken zijn bij de uitvoering van de twee grootschalige onderwijscohorten die momenteel in Nederland lopen – PRIMA en VOCL – en met andere wetenschappers die over deskundigheid beschikken op het terrein van onderwijs en onderwijsonderzoek. Ook zijn gesprekken gevoerd met vertegenwoordigers van de relevante directies bij het ministerie van Onderwijs, Cultuur en Wetenschappen. De verkenning leverde enerzijds belangrijke informatie op over de stand van zaken met betrekking tot longitudinaal onderzoek nu en in het verleden, en anderzijds waardevolle inzichten en gezichtspunten die betrekking hebben op wat op de langere termijn mogelijk en/of wenselijk is. Een lijst van onze gesprekspartners en een overzicht van de door ons geraadpleegde documentatie zijn opgenomen als bijlagen bij dit preadvies.

Tijdens de gesprekken die wij gevoerd hebben kwam al snel naar voren dat we onze opdracht breder moesten opvatten dan oorspronkelijk de bedoeling was. In de eerste plaats werd duidelijk dat door de direct betrokkenen bij PRIMA en VOCL (i.c. onderzoekers, opdrachtgevers, beleid) reeds een uitvoerige evaluatie plaatsvindt over de meest gewenste inrichting van de huidige cohorten¹. We hebben het niet zinvol geacht om een dergelijke evaluatie nog eens dunnetjes over te doen. In de tweede plaats werd duidelijk dat de mogelijkheden om de huidige cohorten op korte termijn op een fundamenteel andere wijze in te richten vrij beperkt zijn. De algemene

1. Zie bijvoorbeeld: *De VOCL-Cohortonderzoeken: Bezinning en Vooruitblik* (1996). Bij PRIMA is in december 1997 een workshop 'Hoe verder met PRIMA?' gehouden.

indruk is dat de cohorten, binnen de huidige randvoorwaarden en gegeven de beleidswensen professioneel en efficiënt zijn opgezet.

Dit laat overigens onverlet dat op de korte termijn wel degelijk op een aantal punten verbeteringen wenselijk zijn. We denken dan met name aan een verbeterde onderlinge afstemming van de cohorten (zowel qua steekproef als qua meetinstrumenten) een verbetering van de toegankelijkheid van de data voor externe onderzoekers (zowel wat betreft het 'toevoegen' van vragen als het gebruik in secundaire analyses) en vermindering van de belasting van scholen door het gebruik van standaardbronnen (met name CITO-toetsen en examengegevens inspectie).

Zoals hierboven opgemerkt, heeft deze studie dus niet tot doel om de bestaande cohorten te evalueren. Veel eerder hebben wij gemeend om als het ware enige afstand te nemen, en de vraag te stellen wat de functie is van cohortonderzoek, en welke plaats deze vorm van onderzoek inneemt in het totale spectrum van onderwijsonderzoek. Cohortonderzoek is geen doel op zich, maar een middel dat kan worden ingezet om inzicht te verschaffen in belangrijke ontwikkelingen en processen in het onderwijs. De uiteindelijke keuze van de onderzoeksopzet zal moeten volgen uit de aard van de inzichten die zijn gewenst.

Hiermee is de tijdshorizon waarop dit preadvies betrekking heeft verschoven van de korte termijn naar de middellange termijn. Het door ons voorgestelde ontwerp verwijst naar de situatie zoals wij ons dat voorstellen op de middellange termijn. Dit staat geheel los van het onderzoek naar de concrete inrichting van de Nederlandse onderwijscohorten op de kortere termijn.

In de volgende paragraaf wordt een overzicht gepresenteerd van de belangrijkste cohortstudies die tot nu toe in Nederland zijn uitgevoerd, aangevuld met een aantal prominente voorbeelden uit het buitenland. Vervolgens worden in het kort de belangrijkste doelstellingen van de twee Nederlandse cohortstudies die momenteel door NWO worden gefinancierd – PRIMA en VOCL – beschreven. In paragraaf 3 schetsen wij een breder onderzoekskader waarbinnen onderwijscohorten kunnen worden geplaatst. Aan de hand hiervan wordt in paragraaf 4 een aantal alternatieve manieren van data vergaren geïnventariseerd, alvorens in paragraaf 5 over te gaan op onze concrete voorstellen voor de middellange termijn. De finale versie van het preadvies hebben wij voorgelegd aan een zestal experts. De belangrijkste punten uit hun commentaar worden samengevat in paragraaf 6.

2 Huidige en vroegere cohorten

Belangrijke cohortstudies in binnen- en buitenland

Als voorbereiding op onze taak hebben wij eerst geprobeerd een beeld te krijgen van het huidige en vroegere cohortonderzoek in Nederland, en ook van studies in het buitenland. Al genoemd zijn de twee grootschalige cohortonderzoeken die in Nederland momenteel lopen op het terrein van het onderwijs, beide door NWO gefinancierd:

- Het PRIMA-cohort in het basisonderwijs. Dit is een doorlopend cohort waarbij elke twee jaar een 'nieuwe' lichting van ongeveer 15.000 leerlingen verspreid over 770 basisscholen wordt gevolgd. De meetmomenten zijn in groep 2, 4, 6 en 8. Door NWO wordt tweejaarlijks een bedrag van 2,5 miljoen gulden beschikbaar gesteld. De totale steekproefomvang bedraagt ongeveer 60.000 leerlingen. Het onderzoek wordt uitgevoerd door ITS en SCO. Het PRIMA-cohort is een voortzetting en uitbreiding van het LEO-cohort dat eind jaren tachtig werd opgezet om het onderwijsvoorrangsbeleid te evalueren. In zowel het LEO-cohort als het PRIMA-cohort worden leerlingen ook ná de basisschool gevolgd in hun schoolloopbaan in het VO en daarna.
- De VOCL-cohorten 1989 en 1993. Deze bestaan uit lichtingen van elk ongeveer 20.000 leerlingen verspreid over 300 scholen voor voortgezet onderwijs. De meetmomenten zijn jaarlijks vanaf klas 1, tot 2 jaar na het schoolverlaten. Er zijn plannen voor het volgen van een nieuw cohort in 1999. Door NWO wordt een maximum budget van 2,1 miljoen gulden op tweejaarbasis gereserveerd. Daarboven investeert het CBS jaarlijks ruim 0,6 miljoen gulden voor de dataverzameling. De uitvoerende instituten zijn het CBS, CITO, GION en OCTO. Voorafgaande aan VOCL zijn door het CBS in 1977 en 1982 twee grootschalige cohortstudies uitgevoerd onder leerlingen van het voortgezet onderwijs (SMVO en SLVO). Deze onderzoeken kunnen op hun beurt weer als vervolg beschouwd worden op het in 1965 gestarte onderzoek 'Van Jaar tot Jaar'. Dit betrof een landelijk representatieve steekproef van ongeveer 3000 leerlingen aan het einde van het basisonderwijs. Deze leerlingen zijn op verschillende tijdstippen weer ondervraagd, doorlopend tot eind jaren zeventig. Het onderzoek werd uitgevoerd door het ITS. Daarnaast kan een aantal regionale cohorten onderscheiden worden, zoals het Noord-Brabant cohort, het Enschede cohort en het Groningen cohort.

In 1987 is het MBO-cohort gestart onder circa 6000 leerlingen van het MTO, MEAO en MDGO. Deze leerlingen zijn ondervraagd ieder jaar totdat ze het MBO hebben verlaten, en ook weer 1 en 3 jaar na afstuderen. Het onderzoek is aanvankelijk door

SVO gefinancierd, en tot de afronding in september 1999 door NWO. Het onderzoek wordt uitgevoerd door GION.

Niet gefinancierd door NWO, maar wel passend binnen de systematiek van de cohortstudies is het zogeheten 'Verder Studeren' panel. Dit bestaat uit een lichting van 2.000 eindexamenkandidaten in 1991 uit de schooltypen VBO, MAVO, HAVO, VWO en MBO, aangevuld met eerste en ouderejaars studenten uit het hoger onderwijs. Dit onderzoek is gefinancierd door het Ministerie van OC&W. De uitvoerende instituten zijn SEO en SCO. Eind 1994 en eind 1995 zijn de deelnemers weer onderzocht.

Ook in het buitenland is in de afgelopen jaren veel cohortonderzoek verricht. Het zou te ver voeren om in deze notitie een overzicht hiervan te presenteren. We volstaan met het kort aanduiden van een aantal prominente studies:

- In de Verenigde Staten is vanaf de jaren '70 een reeks onderzoeken gefinancierd door het *National Center for Education Statistics*. Het eerste hiervan was de *National Longitudinal Study of the High School Class of 1972* (NLS-72), gestart in 1972 onder eindexamenkandidaten uit dat jaar. Dit cohort is weer benaderd in 1973, 1974, 1979 en voor het laatst in 1986. Bij een vergelijkbare cohortstudie, getiteld *High School and Beyond* (HS&B), die in 1980 van start is gegaan, zijn ook leerlingen uit jaar 8 bij de startgroep betrokken. Het cohort is gevolgd tot 1992. De *National Education Longitudinal Study* (NELS) uit 1988 bestaat uitsluitend uit jaar 8 leerlingen uit dat jaar, met vervolgmetingen in 1990, 1992 en 1994.
- Het *Center for Human Resource Research* (CHRR) heeft vanaf midden jaren '60 een serie onderzoeken uitgevoerd in opdracht van het Amerikaanse *Department of Labor*, onder de naam *National Longitudinal Surveys of Labor Market Experience* (NLS). In tegenstelling tot de onderzoeken van het NCES, waar sprake was van uitstroom- of jaargroepcohorten, is hier gekozen voor leeftijdscohorten. Naast de eerste cohorten van mannen en vrouwen van verschillende leeftijdscategorieën die vanaf 1966 zijn gevolgd (eerst jaarlijks, later tweejaarlijks) heeft de in 1979 opgestarte studie de *National Longitudinal Survey of Youth* internationaal veel bekendheid verworven. Dit cohort van mensen die in dat jaar tussen 14 en 21 jaar oud waren, is sindsdien jaarlijks gevolgd. Deze studie kenmerkt zich door de breedte van de vraagstelling, met, naast standaard vragen over arbeidsmarkt ervaringen, in sommige jaren vragen over alcohol- en drugsgebruik, vruchtbaarheid en gezinsvorming.
- Een tweetal Britse studies, de *National Child Development Study* en de *British Cohort Study* (gestart in respectievelijk 1958 en 1970) vertegenwoordigen weer een andere benadering van cohortonderzoek, het geboortecohort. Voor beide

cohorten zijn gegevens verzameld over de (gezins)omstandigheden bij de geboorte en vervolgens de relevante omstandigheden in verschillende latere levensfasen, zoals de vroege ontwikkeling en basisonderwijs, voortgezet onderwijs, beroepsonderwijs, arbeidsmarktintrede en gezinsvorming. Voor beide studies bestond de eerste lichting uit circa 17.000 individuen. Hiervan nemen ongeveer 13.000 mensen nog steeds deel aan beide onderzoeken.

Doelstellingen van PRIMA en VOCL

Voor zowel PRIMA als VOCL geldt dat de keuze voor een cohort-design indertijd is beargumenteerd als een efficiënte manier om met de beperkte financiële middelen die beschikbaar waren een breed scala van doelen na te streven. In de eerste plaats zijn beide cohortstudies opgezet om een aantal beleidsprogramma's te evalueren. Een tweede doelstelling betreft het gebruik van de gegevens voor de beschrijving en registratie van de (verschuivingen in) leerprestaties en schoolloopbaanpatronen, en van de verbanden tussen de kenmerken van de achtergrond en leeromgeving met de prestaties en loopbanen. Een derde doelstelling was om het cohortonderzoek te integreren met het schooleffectiviteitsonderzoek waardoor inzicht wordt verkregen welke kenmerken in de (school-) context bevorderlijk zijn voor de leerprestaties van leerlingen. Dit moest leiden tot een verbetering van de kwaliteitszorg en -bewaking. Tenslotte zouden de cohortonderzoeken worden gebruikt voor wetenschappelijke analyses en toegepast onderzoek voor diverse opdrachtgevers.

Deze multifunctionele opzet is niet alleen efficiënt in financiële termen, maar levert ook op een aantal meer inhoudelijke en onderzoekstechnische punten voordelen op. Door verschillende beleidsprogramma's gelijktijdig te evalueren kan een beter beeld worden verkregen van de netto effecten van de verschillende beleidsprogramma's en de eventuele onderlinge invloeden tussen de programma's. Bovendien worden de scholen niet onnodig vaak lastig gevallen met verzoeken van onderzoekers die elk weer een antwoord op een iets andere vraag trachten te vinden. Tenslotte komt de gekozen opzet de samenhang en de cumulatie van kennis binnen het onderwijs-onderzoek ten goede.

Hoewel deze multifunctionaliteit vanaf het begin bij zowel PRIMA als VOCL werd benadrukt, kan worden vastgesteld dat de cohortstudies grotendeels voortkomen uit de wens tot evaluatie van een aantal grootschalige beleidsoperaties, en ook nooit zouden zijn uitgevoerd zonder de financiële middelen die beschikbaar werden gesteld ten behoeve van deze beleidsevaluaties. Bij PRIMA was dat in eerste instantie de voortzetting van de evaluatie van het Onderwijsvoorrrangsbeleid (OVb), die was begonnen met het LEO-cohort. Daaraan is een aantal andere beleidsprogramma's toegevoegd: Nederlands als Tweede Taal (NT2), Weer Samen Naar School (WSNS), Toerusting en Bereikbaarheid, Sociale Vernieuwing. Hoofddoel bij het VOCL-cohort was de evaluatie van de basisvorming in het voortgezet onderwijs.

In de loop van de tijd is de evaluatie van een aantal andere beleidsprogramma's als doelstelling opgenomen, waaronder OVB, WSNS en het beleid ter zake van Voortijdig Schoolverlaten.

Algemeen kunnen we stellen dat PRIMA en VOCL als doel hebben om zicht te bieden op verschillende aspecten van de onderwijsloopbaan van leerlingen, en in het bijzonder de effecten van het overheidsbeleid hierop. Aandacht wordt bijvoorbeeld besteed aan de vraag hoe leerprestaties worden beïnvloed door kenmerken van de leeromgeving, en hoe kinderen verschillende competenties, houdingen, enz. ontwikkelen. Kortom, ze proberen inzicht te verschaffen in de leerprocessen binnen het onderwijs.

3 'Opbouw, onderhoud en slijtage van menselijk kapitaal'

Vanzelfsprekend zijn onderzoeken naar leerprocessen in het onderwijs in sterke mate gericht op het onderwijs zelf. Zulke leerprocessen maken evenwel deel uit van een veel bredere ontwikkeling die mensen tijdens hun leven doormaken. Leerprocessen in het onderwijs kunnen pas op hun juiste waarde worden ingeschat wanneer die worden gezien tegen de achtergrond van de ontwikkeling die een individu doormaakt tijdens de hele levensloop. Een dergelijke blikverruiming kunnen we in verschillende dimensies zoeken. Ten eerste, kan het schoolse leren zelf worden gezien als één manier waarop mensen hun individuele kwaliteiten (capaciteiten, (meta-)cognities, oriëntaties, enz.) ontwikkelen. Een belangrijk deel van de verwerving van 'menselijk kapitaal' vindt evenwel plaats buiten de directe context van de school, bijvoorbeeld in het gezin of beroep. Ten tweede, beperkt de fase van de opbouw van menselijk kapitaal zich niet tot pakweg de fase van 6-20 jaar (de periode waarin het Nederlandse cohortonderzoek geconcentreerd is), maar bestrijkt eigenlijk het hele leven van mensen. Bovendien geldt zelfs voor de fase van 'schoolse leren' dat veel van de bronnen van leerprocessen in het verleden liggen, en veel van de consequenties ver in de toekomst. Tenslotte, vanuit het perspectief van levenslang leren komt naar voren dat gedurende de levensloop, capaciteiten enz. niet alleen worden opgebouwd, maar ook onderhouden moeten worden en aan slijtage onderhevig zijn. Men kan zich niet eenvoudigweg 'vol laten tanken' op jonge leeftijd en vervolgens het hele verdere leven hierop teren.

Samenvattend kunnen we stellen dat het onderzoek naar schoolloopbanen deel uitmaakt van een breder onderzoeksterrein, dat aan te duiden is met de omschrijving 'opbouw, onderhoud en slijtage van menselijk kapitaal'. Het is wel zinvol om hierbij enige structuur aan te brengen, bijvoorbeeld door verschillende levensfasen te onderscheiden. Elk van deze levensfasen kenmerkt zich door verschillende accenten. Deze accenten hebben betrekking op de ontwikkeling zelf, maar ook bijvoorbeeld op de aard van de keuzes die gemaakt worden en de kenmerken van de omgeving die van belang zijn. De volgende levensfasen kunnen worden onderscheiden met elk hun eigen accenten:

- In de *voorschoolse fase* gaat het om elementaire cognitieve, psychomotorische en sociaal-emotionele ontwikkeling, alsmede om interactieprocessen binnen het gezin die hierop van invloed zijn.
- In het *basisonderwijs* gaat het vooral om de ontwikkeling van schoolse vaardigheden, met als belangrijkste voorspellers de cognitieve en motorische vaardigheden waarmee kinderen het onderwijs binnenstromen, het onderwijs-ondersteunend gedrag vanuit het gezin, en kenmerken van de leeromgeving (i.c. school, klas).

- In het *voortgezet onderwijs* zou het naast prestaties ook moeten gaan om de keuzeprocessen van de leerlingen en de daarbij aansluitende verdere ontwikkeling van kennis en vaardigheden. Bij de te maken keuzen zou het accent niet alleen moeten liggen op de feitelijk gemaakte keuzen, maar ook op het proces zelf. Waarom komt men tot bepaalde keuzen, hoe hangt dit samen met de prestaties, wat is de invloed van de omgeving (gezin, peers, school) etc.? Ten aanzien van de verdere kennisontwikkeling is met name van belang de verwerking van de competenties om door te stromen naar de verschillende vormen van beroeps- en hoger onderwijs.
- In het *beroepsonderwijs en hoger onderwijs* zou het accent moeten liggen op competentie-ontwikkeling en de ontwikkeling van algemene analytische vaardigheden (meta-cognities): op welke wijze kunnen individuen hun competenties op een bepaalde terrein optimaal ontwikkelen en hoe hangt dit samen met kenmerken van de leeromgeving (leerweg, studeerbaarheid etc.)?
- Bij de *transitie van school naar werk* zou het accent vooral moeten liggen op zaken als kansen op werk, aansluiting met de gevolgde opleiding en de arbeidsmarkt- en loopbaanoriëntatie.
- Bij de *beroepsloopbaan* gaat het vooral om opbouw, onderhoud en slijtage van kennis in relatie tot het maatschappelijk functioneren in het algemeen en het beroepsmatig functioneren in het bijzonder en de rol van scholing daarbij.

In schema:

Accenten			
Periode	cognitieve ontwikkeling	keuzes	omgeving
voorschools	elementaire cognities		gezin
basisonderwijs	schoolse vaardigheden	schoolkeuze	gezin, leeromgeving
voortgezet onderwijs	studievaardigheid, zelfstandigheid	leerroutes, studien en beroepskeuze	gezin, peers, leeromgeving
beroepsonderwijs/ hoger onderwijs	competenties, analytisch denken	specialisatie/ generiek of specifiek	leeromgeving
transitie van school naar werk	arbeidsmarktorientatie, loopbaanvaardigheden	arbeidsmarkt-positie	arbeidsmarkt
maatschappelijke carrière en beroepsloopbaan	maatschappelijk en i.h.b. beroepsmatig functioneren/ loopbaanontwikkeling, opbouw/onderhoud/ slijtage van kennis	baankeuze(n), scholing, gezinsvorming	arbeidsorganisatie(s), gezin en andere maatschappelijke domeinen

Het door ons geschetste onderzoekskader heeft belangrijke consequenties voor hoe het onderzoek naar leerprocessen in het onderwijs wordt ingericht. Enerzijds moet het onderzoek voldoende recht kunnen doen aan de accenten in de desbetreffende periode (onderwijsfase). Anderzijds is het duidelijk dat de manier waarop de gemaakte keuzes en de omgevingskenmerken de ontwikkeling van kennis en vaardigheden in een bepaalde periode beïnvloeden, ook medebepalend zal zijn voor de ontwikkeling in de daarop volgende perioden. Er is daarom behoefte aan onderzoek waarmee niet alleen het patroon van causale relaties binnen de onderwijsloopbaan kan worden onderzocht, maar ook de causale relaties tussen wat in het onderwijs gebeurt en de perioden daarvóór en daarna. Eveneens van belang is het om zicht te krijgen op belangrijke transities, zowel binnen de onderwijsloopbaan als tussen de voorschoolse periode en het basisonderwijs en bij het schoolverlaten en de intrede op de arbeidsmarkt.

Vanwege de longitudinale opzet van cohortonderzoek lijkt het voor de hand te liggen dat deze vorm van onderzoek hierin een belangrijke rol zal spelen. Echter, om de waarde en betekenis van cohortonderzoek goed te kunnen inschatten, is het zinvol om ook andere mogelijke onderzoeksvormen te bekijken. In de volgende paragraaf passeren een aantal onderzoeksvormen de revue, waarbij de bruikbaarheid van de besproken benaderingen voor het onderzoek naar de opbouw en onderhoud van menselijk kapitaal centraal staat.

4 De (on)mogelijkheden van cohortonderzoek in vergelijking tot andere databronnen

Alvorens in te gaan op de mogelijkheden van, maar ook de problemen met panel-data op het onderzoeksterrein van 'menselijk kapitaal', zullen we eerst kort een tweetal mogelijke alternatieve databronnen voor grootschalig kwantitatief onderzoek bespreken, namelijk cross-sectioneel onderzoek en registratiebestanden.²

De mogelijkheden om informatie over de opbouw en onderhoud van menselijk kapitaal te verkrijgen door middel van een retrospectieve bevraging in cross-sectioneel onderzoek zijn wel aanwezig, maar kennen een aantal beperkingen. Cross-sectioneel onderzoek lijkt met name geschikt om de feitelijke schoolloopbaan in kaart te brengen, althans waar het om betrekkelijk 'harde' feiten gaat. Waar het gaat om 'zachte' loopbaaninformatie, zoals de ontwikkeling van schoolse vaardigheden of vaktechnische kennis, zijn de mogelijkheden van cross-sectioneel onderzoek echter beperkt. Ook is het design minder krachtig waar het gaat om de mogelijkheid om patronen van causaliteit vast te stellen. Dit geldt vooral voor effecten die zich op de langere termijn afspelen.

Registraties zoals het toekomstig in te voeren onderwijsnummer, zijn potentieel zeer interessant voor het onderzoek naar de ontwikkelingen bij individuen, om twee redenen. Ten eerste, ze bevatten informatie over leerlingen over hun hele onderwijsloopbaan. Ten tweede, ze omvatten (nagenoeg) de hele populatie van leerlingen, waardoor representativiteit ten opzichte van ieder denkbaar referentiekader (leeftijd, instroom, uitstroom) geen enkel probleem vormt. In de praktijk zijn dergelijke registraties echter per definitie beperkt van opzet, en bevatten in de regel slechts enkele 'harde' kerngegevens. Hiermee kunnen in feite alleen globale leerlingstromen in kaart worden gebracht. Ze bieden geen zicht op de cognitieve ontwikkeling van leerlingen of de rol van de omgeving daarbij.

Vergeleken bij de genoemde onderzoeksvormen lijkt cohortonderzoek heel aantrekkelijk. Cohortonderzoeken kenmerken zich door het feit dat er meerdere metingen plaatsvinden op verschillende tijdstippen binnen eenzelfde steekproef van individuen. Algemeen gezien geven cohortgegevens vanwege de longitudinale opzet zicht op de volgorde van gebeurtenissen, en op de veranderingen die over de tijd optreden bij individuen. Hiermee is cohortonderzoek bij uitstek geschikt voor het blootleggen van causale relaties, bijvoorbeeld tussen gedrag of disposities van individuen op tijdstip $t=1$ en hun toestand op een later tijdstip, of tussen omgevings-

2. In het onderwijsonderzoek komen ook andere databronnen voor, zoals experimenteel onderzoek of case-studies. Deze laten zich echter door hun heel andere opzet moeilijk vergelijken met het cohortonderzoek.

invloeden en gedrag. Dergelijke gegevens leveren daarom een krachtig instrument voor wetenschappelijk onderzoek, voor de evaluatie van beleid, en voor de terugkoppeling naar de dagelijkse (onderwijs)praktijk. Daarnaast kunnen cohort-onderzoeken een belangrijke monitoringfunctie vervullen, waardoor belangrijke maatschappelijk tendensen zichtbaar kunnen worden gemaakt en problemen en knelpunten sneller kunnen worden gesignaleerd.

Zoals blijkt uit het overzicht in paragraaf 2, kenmerken cohortonderzoeken zich evenwel door een veelheid aan verschijningsvormen. Zowel in termen van levensfase als van maatschappelijk domein kunnen ze heel algemeen of tamelijk specifiek zijn. Een heel algemene vorm van cohortonderzoek die ook relatief veel wordt toegepast, betreft de meting van individuen over (vrijwel) de hele levensloop. Voorbeelden van zo'n benadering zijn de eerder genoemde Britse geboortecohortstudies. Dergelijke levenslange metingen lijken bij uitstek geschikt om effecten op lange termijn te onderzoeken. Echter, hoewel zo'n lange-termijn cohort voor veel vraagstellingen nuttige informatie kan opleveren, zijn er belangrijke problemen die de bruikbaarheid van de opgeleverde informatie reduceren. Een probleem is dat een levenslange meting op de korte termijn slechts informatie oplevert over heel jonge kinderen. Een dergelijk onderzoek moet al een lange tijd lopen voordat het informatie oplevert over individuen in latere levensfasen. Een ernstiger probleem is het feit dat zulke studies zich meestal door een hoge uitval kenmerken. Het is daarom moeilijk om een representatieve steekproef te handhaven voor onderzoek naar de latere levensfasen. Verder is het moeilijk om in zo'n brede opzet recht te doen aan veel van de vragen die voor onderwijsonderzoek van belang zijn.

Periode-specifieke onderwijscohorten (zoals PRIMA en VOCL) zijn daarentegen te bestempelen als vrij specifiek, zowel in termen van levensfase als van maatschappelijk domein. Gezien de belangrijke rol die aan het onderwijs wordt toegekend in de ontwikkeling van individuele capaciteiten, en daarmee in het bepalen van levenskansen, is de belangstelling voor dit soort onderzoek in het onderwijs niet vreemd. Dit geldt des te meer omdat het onderwijs – tot op zekere hoogte – vatbaar is voor beïnvloeding door beleid. Maar, zoals in de vorige paragraaf werd opgemerkt, wat in het onderwijs gebeurt staat niet op zichzelf. Zo wordt de ontwikkeling van bijvoorbeeld competenties en houdingen tijdens de schoolloopbaan mede bepaald door de (vroege) gezinssituatie. Op hun beurt hebben deze competenties en houdingen een sterk effect op het verdere verloop van het leven van individuen: het (eventueel) vervolgonderwijs, de beroepsloopbaan, en het maatschappelijke functioneren in brede zin. De sterkste effecten zijn hierbij wellicht te verwachten tijdens perioden van grote veranderingen, en met name bij de transities tussen de verschillende levensfasen (zie schema). Maar ook de lange termijn effecten zijn van belang. Cohorten zoals PRIMA en VOCL die zich primair op een bepaalde fase in het onderwijs richten zullen met name deze lange-termijn effecten voor een belangrijk deel missen.

Tot op zekere hoogte zijn deze problemen op te vangen door een cohort dat wordt gevolgd in een bepaalde onderwijsfase, voor een deel te blijven volgen tijdens een volgende onderwijsfase, of nadat de respondenten het onderwijs hebben verlaten. Dit gebeurt al tot op zekere hoogte in de huidige onderwijscohortstudies. Ook het weer wakker maken van 'slapende' cohorten (cohorten die eerder zijn onderzocht en waarvan de adresgegevens goed zijn bewaard en geactualiseerd) zou een zeer goede optie kunnen zijn. In principe kan zo een combinatie van cohorten uit verschillende fasen een 'dakpanconstructie' vormen waarmee het totaalbeeld kan worden gereconstrueerd. Echter, in de praktijk is de bruikbaarheid van gegevens die op een dergelijke manier zijn vergaard tamelijk beperkt, tenzij men deze gegevens kan ophangen aan een deugdelijk populatiebestand. De reden is dat het moeilijk is om de resultaten uit verschillende onderzoeken die zich primair op één onderwijsfase richten met elkaar in verband te brengen wanneer de relatie van de gebruikte steekproeven tot de totale populatie niet bekend is.

De representativiteit van een cohort kan om twee redenen in gevaar worden gebracht. Ten eerste is er het probleem van selectieve uitval of panelmortaliteit. Zelfs bij geringe uitval per meting kan het cumulatieve effect hiervan op het totale design erg groot zijn. Het probleem is niet alleen dat uitval plaatsvindt, maar ook dat we vaak niet weten of deze op relevante kenmerken (bijvoorbeeld de schoolloopbaan) selectief van aard is. Wanneer er sprake is van selectiviteit, en wanneer de aard en omvang hiervan niet goed te achterhalen is, kunnen zelfs vergelijkingen tussen verschillende lichtingen van eenzelfde cohortstudie zeer problematisch zijn.

Een tweede probleem heeft te maken met de wijze waarop de steekproef getrokken wordt. Idealiter wil men in cohortonderzoek een bepaalde leeftijdsgroep volgen. In plaats hiervan wordt in het onderwijsonderzoek vaak gekozen voor het volgen van een instroomcohort. Dit levert een zekere selectiviteitsbias op ten opzichte van de relevante leeftijdsgroep in de populatie. Door veranderingen in beleid zijn bepaalde categorieën leerlingen – die nominaal hetzelfde zijn – in sommige gevallen feitelijk niet meer goed te vergelijken. Dit betekent dat een deel van de beoogde beleids-effecten niet zichtbaar wordt, hetgeen een adequate evaluatie bemoeilijkt³.

Ook de keuze, veelal uit praktische overwegingen, om steekproeven van scholen te trekken en daarbinnen de leerlingen te volgen levert een selectiviteitsbias op, wanneer de betreffende leerlingen uitwaaieren naar andere scholen die niet in het

3. Een voorbeeld kan dit duidelijk maken. Stel we vergelijken twee instroom cohorten ($t=1$ en $t=2$) in het hoger onderwijs. Wanneer tussen $t=1$ en $t=2$ de deelname aan hoger onderwijs sterk is gestegen betekent dit dat de samenstelling van beide instroomcohorten is veranderd. Voor een beoordeling van verschuivingen in bijvoorbeeld het studierendement van beide cohorten is het nodig om niet alleen de cohorten zelf te vergelijken, maar ook om deze te relateren aan de desbetreffende totale leeftijdspopulatie.

onderzoek betrokken zijn. Area-sampling kan de ernst van dit probleem weliswaar verminderen, maar doet aan het principiële probleem niets af dat bij cohortonderzoek het primaat bij het volgen van leerlingen ligt.

Bovengenoemde problemen zijn weliswaar niet op te lossen, maar wel beter in de hand te houden, wanneer de betreffende cohortgegevens gekoppeld kunnen worden aan een populatiebestand. In dit verband zijn momenteel juist enkele ontwikkelingen gaande die uitzicht bieden op geheel nieuwe mogelijkheden voor onderzoek in het onderwijs. Het gaat hier enerzijds om een aantal technische innovaties, die de mogelijkheid creëren om op een veel efficiëntere manier informatie te vergaren en uit te wisselen. Anderzijds gaat het om een verandering in de wetgeving, in de vorm van de invoering van het integraal onderwijsnummer. Met het oog op deze mogelijkheden schetsen wij in paragraaf 5 een opzet voor een mogelijke data-infrastructuur voor het longitudinale onderzoek in de toekomst.

5 De data-infrastructuur in de toekomst

Periodespecifieke cohortonderzoeken zitten vaak in een spanningsveld tussen enerzijds het recht willen doen aan de desbetreffende onderwijsperiode, en anderzijds het willen benutten van de voordelen van een longitudinale opzet. Uiteindelijk kiezen onderzoekers noodgedwongen voor een compromis waarin geen van beide aspecten goed tot hun recht komen. Omdat het mogelijk moet zijn de steekproef over een aantal jaren te kunnen blijven volgen, wordt voor een steekproefopzet gekozen waarvan selectiviteit ten opzichte van de desbetreffende leeftijdsgroep op landelijk niveau niet uit te sluiten is. Aan de andere kant, wordt het onderzoek in de kern beperkt tot slechts één periode, omdat de onderzoekers specifieke vragen willen kunnen beantwoorden over deze onderwijsperiode. De lange-termijn effecten van het onderwijs blijven hierdoor grotendeels buiten beeld.

Om uit dit spanningsveld te kunnen ontsnappen is op de middellange termijn de meest gewenste opzet een systeem waarbij verschillende databronnen worden gecombineerd. We zouden dan ook niet meer willen spreken van een 'onderzoek' maar van een 'data-infrastructuur'. De basis hiervan zou moeten worden gevormd door een landelijk populatiebestand waarbij de belangrijkste kerngegevens van alle leerlingen in het onderwijs worden geregistreerd. Dit bestand zou gegevens moeten bevatten over sociale achtergrond en onderwijsposities, maar ook bijvoorbeeld CITO-toetsresultaten en andere prestatie-indicatoren. In principe moet aan de hand van deze gegevens in grote lijnen de hele schoolloopbaan van iedere Nederlander – in termen van 'harde' indicatoren – te reconstrueren zijn. Het lijkt ons wenselijk dat deze infrastructuur breed van opzet is, en wordt gebruikt en onderhouden door diverse belanghebbende organisaties, niet alleen, of zelfs in de eerste plaats, door organisaties die zich bezighouden met onderwijsonderzoek.

Om de nodige diepte te kunnen bereiken moet deze basisregistratie worden aangevuld met cross-sectioneel en longitudinale deelonderzoeken van specifieke ontwikkelingsfasen en/of van specifieke deelpopulaties. Het integraal basisbestand vormt hiervoor een landelijk steekproefkader. Omdat de 'harde' gegevens beschikbaar zijn via de basisregistratie, kunnen deze deelstudies zich concentreren op specifieke ontwikkelingsvariabelen, zoals keuzeprocessen, ontwikkeling van meta-cognities, en specifieke competenties. Deze aanvullende deelonderzoeken zouden zich niet moeten beperken tot de periode waarin leerlingen op school zitten, maar ook de voorschoolse periode en de verdere levensloop moeten bestrijken. Voorzover dit buiten het directe aandachtsgebied van de PROO komt te liggen, betekent dit dat ook andere financiers bij de opzet van een dergelijke data-infrastructuur betrokken moeten worden. Ook gegevens uit cross-sectionele surveys op een hoger aggregatieniveau (klas, school, stad, regio) zouden gekoppeld moeten kunnen worden aan de basisregistratie op individueel niveau.

Door de koppeling aan een basisregistratie kunnen de aanvullende deelonderzoeken veel beperkter en efficiënter zijn qua opzet. Ook is het niet strikt nodig om deelpopulaties altijd langdurig te volgen, omdat de informatie uit verschillende cohortonderzoeken door de koppeling met de basisregistratie makkelijker met elkaar in verband is te brengen, bijvoorbeeld door de toepassing van imputatietechnieken.

Het schema Opzet data-infrastructuur geeft een overzicht van de verschillende informatiebronnen in het door ons voorgestelde ontwerp. Het informatiesysteem bestaat uit vijf delen. De gegevens in de eerste twee delen vormen samen het populatiebestand. Deze gegevens worden in principe bijgehouden voor alle individuen over (vrijwel) de hele levensloop. Het eerste deel verwijst naar de basisregistratie met een uniek identificatienummer per individu, en andere 'duurzame' kerninformatie zoals het ouderlijk milieu, het geslacht en de geboortedatum. Het waarschijnlijk binnenkort in te voeren persoonsgebonden onderwijsnummer zou fungeren als identificatienummer. Alle overige gegevens in het informatiesysteem worden aan het onderwijsnummer gekoppeld, eventueel via een vertrouwelijke sleutel. De verantwoordelijkheid voor het beheer van het onderwijsnummer ligt bij het Ministerie van OC&W.

Het tweede deel betreft longitudinale 'kerninformatie' over alle individuen in het bestand. In tegenstelling tot de eerste categorie gaat het hier om periode-specifieke informatie, dat wil zeggen gegevens die kunnen veranderen over de tijd, zoals onderwijsposities en prestatie-indicatoren op school. Tijdens de onderwijsloopbaan zijn de belangrijkste informatiebronnen hierbij de onderwijsinstellingen zelf en CITO, danwel onderwijsinspectie of IBG. Door deze gegevens krijgt het populatiebestand over de levensloop een cumulatief karakter.

Aanvullingen op dit populatiebestand worden vooral geleverd door onderzoeksinstituten. In het algemeen gaat het om vormen van onderzoek zoals nu al wordt uitgevoerd. Voorbeelden van cross-sectionele aanvullingen zijn de Enquête Bezoeksbevolking van het CBS en de schoolverlatersenquête van het ROA. Wat betreft longitudinale studies zouden onderzoeken zoals PRIMA en VOCL deel uitmaken van het informatiesysteem. Deze cohorten zouden dan veel eenvoudiger en efficiënter van opzet kunnen worden, omdat een deel van de oorspronkelijke doelstelling (de beschrijving van schoolloopbanen op basis van enkele harde kern-indicatoren) kan komen te vervallen. Het is eveneens denkbaar dat geboortecohorten, vergelijkbaar met bovengenoemde Britse studies, aan het populatiebestand worden gekoppeld. Tenslotte kunnen context-gegevens, die betrekking hebben op eenheden op een hoger aggregatieniveau, via de basisregistratie aan de andere bestanden worden gekoppeld. Voor al deze studies zou het integraal populatiebestand een landelijk steekproefkader vormen.

Opzet data-infrastructuur

onderdeel:		soort informatie:	gegevens vergaard/geleverd door:
populatiebestand	basisregistratie, 'duurzame' kerninformatie	onderwijsnummer, achtergrondgegevens	ministerie van OC&W, onderwijsinstellingen
	periode-specifieke kerninformatie	BRIN-nummer, onderwijsposities, prestatie-indicatoren	CITO, onderwijsinspectie, IBG, onderwijsinstellingen
aanvullend onderzoek	cross-sectioneel	diepte-informatie op basis van steekproef onder (deel van) populatie op één tijdstip	CBS / onderzoeksinstituten
	longitudinaal	diepte-informatie op basis van steekproef onder (deel van) populatie over één of meerdere levensfasen	CBS / onderzoeksinstituten
	context-informatie	informatie over scholen, schoolklassen,	CBS/onderzoeksinstituten

Om een dergelijke opzet daadwerkelijk te kunnen implementeren, moet aan een aantal basisvoorwaarden worden voldaan. De meest essentiële vereiste is het bestaan van een uniek identificatienummer om de gegevens aan op te hangen, en om de koppeling met andere bestanden in principe mogelijk te maken. In dit verband is het waarschijnlijk binnenkort in te voeren persoonsgebonden onderwijsnummer van grote betekenis. Het wetsontwerp betreffende de invoering van het onderwijsnummer is op 22 december 1997 bij de Tweede Kamer ingediend. Het is de bedoeling om als onderwijsnummer het Sofi-nummer te gebruiken. De verwachting is dat het wetsontwerp in de loop van 1999 door de Eerste en Tweede kamer zal worden goedgekeurd. Vervolgens zal het waarschijnlijk nog een paar jaar duren voordat de gehele basisadministratie over de volle breedte operationeel is.

Een tweede voorwaarde is dat in dit registratiebestand alle centrale kernvariabelen aanwezig zijn, en zodanig zijn geoperationaliseerd dat ze een reconstructie van de schoolloopbaan mogelijk maken. Vooralsnog wordt aan het onderwijsnummer slechts een beperkt aantal gegevens betreffende de sociale achtergrond en onderwijspositie van leerlingen gekoppeld. Op zijn minst moet het bestand worden aangevuld met essentiële informatie over de ontwikkeling in prestaties zoals CITO-toetsresultaten en eindexamengegevens. In beginsel is een groot deel van deze informatie (CITO-toetsen, examencijfers) reeds beschikbaar. In het licht van de maatschappelijke discussie over kwaliteitsbewaking en openbaarheid van onderwijsprestaties lijkt de trend ook te zijn dat op termijn een uniforme verplichting zal gaan gelden voor onderwijsinstellingen om dit type gegevens te leveren.

Een derde voorwaarde is dat het mogelijk moet zijn voor onderzoekers om gebruik te maken van het onderwijsnummer en de daarbij behorende gegevens, alsmede om het onderwijsnummer en de bijbehorende gegevens te koppelen aan informatie uit allerlei soorten aanvullend onderzoek. Het wetsontwerp biedt hiervoor wel enige mogelijkheden, maar uiteraard zullen afdoende maatregelen getroffen moeten worden om de privacy van individuen te waarborgen. In dat verband is het zinvol om te verwijzen naar de situatie in Denemarken waar een vergelijkbare infrastructuur reeds lang functioneert⁴.

Indien zowel de Tweede als de Eerste Kamer het wetsontwerp goedkeuren, zal daarmee aan de eerste voorwaarde zijn voldaan. Vooral wat betreft de andere voorwaarden heerst op dit moment een hoge mate van onzekerheid. Ondanks het feit dat bij de invoering van het onderwijsnummer het verbeteren van de kwaliteit en beschikbaarheid van beleidsinformatie een van de hoofddoelstellingen is – naast het vergroten van de controle op de onderwijsbestedingen en het verminderen van de administratieve belasting van de scholen – lijkt men ervan uit te gaan dat de op-

4. Zie Eurostat, *Statistics on Persons in Denmark a register-based Statistical System*, 1991, ISBN 22-827-4005-6.

brengst van het onderwijsnummer voor onderzoek ten behoeve van beleid en wetenschap tamelijk bescheiden zal zijn. Naar onze mening is deze terughoudendheid misplaatst. Sterker nog, wil men de kwaliteit van het onderwijs behoorlijk kunnen blijven onderzoeken, dan lijken er op termijn weinig reële alternatieven te zijn.

Wanneer het doel van het onderzoek is om inzicht te verwerven in het proces van opbouw, onderhoud en slijtage van menselijk kapitaal, is het van belang dat de juiste informatie wordt gevraagd bij de juiste groep personen op de relevante tijdstippen. In feite zorgen de organisatorische eisen van de cohortstudies er meestal voor dat de opzet van de dataverzameling vooral een kwestie is van wat haalbaar is, gelet op de restricties waarbinnen dit moet gebeuren. De restricties zijn doorgaans niet gering, en lijken alleen maar toe te nemen met de tijd. Van de zijde van de opdrachtgevers bestaat een grote pressie die leidt tot een opvoering van de veelheid en complexiteit van de vragen die met de cohortgegevens zouden moeten worden beantwoord. De ontwikkelingen binnen het onderwijs maken het steeds moeilijker om een representatieve steekproef in stand te houden. Leerlingen waaiëren uit over geheel verschillende onderwijsloopbanen. Schoolfusies vooral bij het voorgezet onderwijs leveren grote problemen voor de onderzoekers. Bovendien neemt de belasting van scholen in termen van verzoeken om mee te doen aan onderzoek sterk toe. De bereidheid van scholen om mee te blijven werken kan alleen worden gewaarborgd wanneer er strikte afspraken worden gemaakt om individuele scholen niet onevenredig te belasten. Onderwijscohorten dreigen daarmee in een impasse te raken tussen de kennisdoelen en de organisatorische restricties.

Het bovengeschetste ontwerp biedt in principe een uitweg uit deze impasse. Door informatie uit verschillende bronnen te koppelen aan eenzelfde populatiebestand, kan de opzet van de onderzoeken in sterke mate worden toegespitst op de gestelde vragen, zonder dat de representativiteit van de gegevens een probleem hoeft te worden. Door gebruik te kunnen maken van de kerninformatie in het populatiebestand kunnen vragenlijsten aanzienlijk worden ingekort en kan ook de cohortomvang sterk worden beperkt. Hierdoor krijgt tevens ieder afzonderlijk onderzoek in principe een lange-termijn karakter, omdat men over essentiële informatie vóór en ná de eigenlijke meting kan beschikken. Ook het weer 'wakker maken' van 'slappende' cohorten behoort altijd tot de mogelijkheden.

Over de nadere vormgeving – met name organisatorisch – van een dergelijke data-infrastructuur valt uiteraard veel meer te zeggen. We hebben het echter niet tot onze taak gerekend hier een complete blauwdruk te leveren voor een ideale organisatorische opzet. We hebben slechts het belang willen schetsen van een infrastructuur waarbinnen verschillende databronnen kunnen worden gecombineerd, en de belangrijkste beslispunten voor de realisering daarvan willen aangeven. Naar onze

mening heeft het pas zin om de organisatorische en financiële details verder uit te werken als het duidelijk is dat er voldoende draagvlak is voor het idee.

Wanneer men de noodzaak van een dergelijk informatiesysteem onderschrijft, zal ervoor moeten worden gezorgd dat aan alle drie bovengenoemde voorwaarden kan worden voldaan. Wanneer het onderwijsnummer wordt ingevoerd, moet het mogelijk worden gemaakt om CITO-toets scores en andere prestatie-indicatoren gedurende de schoolloopbaan hieraan te koppelen. Het resulterende populatiebestand zou in principe toegankelijk moeten zijn voor gebruik voor legitieme onderzoeksdoelen, waarbij het mogelijk moet zijn om informatie uit aanvullend onderzoek aan het populatiebestand te koppelen.

Als het lijkt dat aan de voorwaarden kan worden voldaan, kan een stappenplan worden ontwikkeld voor het verwezenlijken van de voorgestelde data-infrastructuur. Vervolgens zou elke stap in detail moeten worden uitgewerkt in nauw overleg met deskundigen op het desbetreffend gebied.

De resulterende data-infrastructuur zou kunnen worden gebruikt voor een breed scala van doelen, zonder dat concessies hoeven te worden gedaan ten aanzien van de kwaliteit van de gegevens. Enerzijds zou het bruikbare effectmetingen van beleid en andere onderwijsveranderingen toelaten, waarbij op meerdere tijdstippen onderwijsgegevens worden geanalyseerd over een breed aantal kenmerken. Anderzijds biedt deze opzet de mogelijkheid voor diepte analyses over de ontwikkeling van menselijk kapitaal van de wieg tot het graf, waarbij echt naar de (relevante) lange-termijn processen en gebeurtenissen kan worden gekeken.

6 Commentaar van experts op het preadvies: hoofdpunten

De finale versie van het preadvies is gestuurd aan zes experts, met het verzoek om commentaar. Deze waren: Prof. dr. H. Ganzeboom (UU), Prof. dr. J. Hartog (UvA), Prof. dr. J. Dronkers (SCO/UvA), Prof. dr. J. Peschar (RUG), dr. C. van Bochove (CBS) en Prof. dr. P. Schnabel (SCP). Schriftelijke reacties werden ontvangen van Ganzeboom, Hartog, Dronkers, Peschar, Van Bochove en (namens het SCP) drs. T. Roes. De belangrijkste punten uit deze reacties zijn hieronder kort samengevat. De reacties zelf zijn integraal afgedrukt in bijlage 3.

Alle experts onderschrijven het idee van een data-infrastructuur waarbij gegevens uit verschillende bronnen aan elkaar worden gekoppeld via een volledige registratie. Een dergelijke koppeling van gegevens uit verschillende bronnen bespaart tijd en geld door onnodige overlap te elimineren, ontlast daarbij onderzoekers en scholen, bevordert een ver gaande coördinatie tussen verschillende onderzoeken, en leidt tot een aanzienlijke verhoging van de informatiewaarde van de verzamelde gegevens.

Niettemin zetten de experts kanttekeningen bij het door ons voorgesteld ontwerp. Het commentaar concentreert zich vooral op vier belangrijke issues: privacy, toegankelijkheid, het gebruik van het onderwijsnummer, en de organisatorische opzet van de data-infrastructuur.

Alle experts benadrukken het belang van een goed systeem van beveiliging voor de data-infrastructuur, om ervoor te zorgen dat de privacy van individuele leerlingen niet wordt geschonden. Niemand lijkt er van uit te gaan dat dit als zodanig een onoverkomelijke probleem is. Er zijn voorbeelden in andere landen, zoals Denemarken en Zweden, die laten zien dat er goed werkende oplossingen voor dit probleem te vinden zijn. Verder wordt opgemerkt dat ook in Nederland de animo om naar oplossingen te zoeken voor problemen rondom privacy in recente jaren sterk is toegenomen.

Een aantal experts is van mening dat dergelijke privacy overwegingen niet ten koste mogen gaan van een ruim gebruik van de gegevens. Volgens deze experts is het van groot belang dat toegang tot de gegevens niet wordt beperkt tot instellingen die dicht bij de overheid staan. Onafhankelijke onderzoekers zouden dus actief moeten worden gestimuleerd om de gegevens te gebruiken. Verder zouden geen onnodige beperkingen moeten worden gelegd op de informatie waarover men zou mogen publiceren. In dit verband wordt opgemerkt dat het in Denemarken niet is toegestaan om over individuele scholen te rapporten.

Hoewel de principiële voordelen van het gebruik van het onderwijsnummer door alle experts worden ingezien, heeft een aantal van de experts enige reserves over de praktische implicaties van het gebruik daarvan. Er wordt gewezen op de enorme inspanning die vereist zou zijn om de nodige adressenmutaties aan te brengen. Dit probleem is voor een groot deel op te lossen door een koppeling met de Gemeentelijke Basisadministratie (GBA), maar dit zou tot additionele bezwaren in verband met de privacy-wetgeving kunnen leiden. Verder wordt gevreesd dat de politieke gevoeligheid van het onderwijsnummer die voortvloeit uit zulke privacy-overwegingen de bruikbaarheid daarvan zou kunnen aantasten. Deze gevoeligheid zou kunnen worden gebruikt om concessies ten aanzien van de toegankelijkheid van de gegevens te rechtvaardigen. Om dezelfde redenen worden problemen voorzien waar het gaat om het aanvullen van de gegevens in de basisregistratie met goede indicatoren van de sociaal-culturele achtergrond. De achtergrondgegevens behorend bij het onderwijsnummer zouden hiervoor volstrekt ontoereikend zijn, vooral waar het om zaken als etniciteit gaat. Juist zulke gegevens zijn extra gevoelig uit oogpunt van privacy.

Bovenstaande punten hangen samen met verschillende visies omtrent de organisatorische opzet van de data-infrastructuur. Een belangrijke kritiekpunt op het ontwerp is dat er geen concrete uitspraken worden gedaan ten aanzien van de vraag wie de data-infrastructuur zou moeten beheren en hoe zaken zoals privacy en toegankelijkheid van de gegevens formeel moeten worden geregeld. Als potentiële beheerders worden NWO, ministeries, CBS, SCP en KNAW genoemd. Een aantal experts draagt hierbij argumenten aan om de voorkeur voor bepaalde instituten te onderbouwen.

Waar het om het op korte termijn opzetten en beheren van een dergelijke data-infrastructuur gaat, wordt het CBS voorgesteld als een zeer geschikte kandidaat. De Wet op het CBS en de Centrale Commissie voor de Statistiek regelt dat het CBS gebruik mag maken van alle registraties van alle rijksinstellingen en -diensten. Bovendien beschikt het CBS nu al over technische en organisatorische voorzieningen om de geheimhouding te waarborgen. Hiertegen wordt door sommige experts ingebracht dat wanneer het beheer bij een instituut als het CBS zou liggen, de toegankelijkheid van de gegevens voor onafhankelijke onderzoekers in gevaar zou komen. De terbeschikkingstelling van gegevens voor legitieme onderzoeksdoeleinden zou volgens deze visie het beste kunnen worden gewaarborgd wanneer het beheer bij een instelling als NWO of KNAW zou liggen. Als onderdeel van de data-infrastructuur zou daarbij een geormerkte geldstroom moeten komen voor analyses met en aanvullingen op de bestanden.

Naast deze algemene discussiepunten is ook een aantal meer specifieke suggesties en opmerkingen naar voren gebracht. Er wordt op gewezen dat de kosten en baten van het gebruik van een volledige registratie nog altijd dienen te worden vergeleken

met mogelijke alternatieven zoals het jaarlijks benaderen van een groot cohort. Om voldoende greep te krijgen op de verschillende manieren waarop de factor tijd een rol kan spelen (i.c. cohort-, periode- en levensfase-effecten), wordt het van essentieel belang geacht dat er verschillende leeftijdscohorten gelijktijdig en op vergelijkbare wijze worden gevolgd. Verder wordt opgemerkt dat het geschetste ontwerp betrekking heeft op de lange termijn. Het is derhalve raadzaam om maatregelen te treffen om de bruikbaarheid op de korte termijn te garanderen. Het zou bijvoorbeeld mogelijk zijn om aanvankelijk aan de hand van retrospectieve bevraging zicht te krijgen op eerdere levensfasen, waarbij deze gegevens geleidelijk aan kunnen worden vervangen door echte longitudinale gegevens. Er zou ook kunnen worden nagegaan of er teruggegrepen kan worden op bestaand materiaal, zoals CITO toetsresultaten uit de afgelopen decennia. Een andere expert merkt echter op dat het CITO tot nu toe de toetsresultaten erg streng bewaakt, en vraagt zich af of CITO er mee akkoord zal gaan dat CITO-scores aan de basisregistratie worden gekoppeld.

Tenslotte wordt opgemerkt dat er andere concepten denkbaar zijn die een afbakening van het onderzoeksterrein kunnen geven. Buiten het in het voorstel genoemde 'proces van opbouw, onderhoud en slijtage van menselijk kapitaal' worden concepten als 'reproductie en verandering van sociale ongelijkheid' en 'sociale integratie en -uitsluiting' als mogelijke verbreding van het onderzoeksterrein genoemd.

Bijlage 1: Lijst van gevoerde gesprekken

Datum	Expert(s):	Instituut:	Aanwezig van het ROA:
27-6-1997	Prof. dr. H. Ganzeboom	UU	Van der Velden
6-11-1997	Prof. dr. J. Dronkers Prof. dr. W. Meijnen	SCO SCO	Allen, Heijke, Van der Velden
6-11-1997	dr. H. Oosterbeek dr. U. de Jong	UvA SCO	Allen, Heijke, Van der Velden
14-11-1997	dr. P. Jungbluth	ITS	Allen
14-11-1997	dr. R. Bosker	OCTO	Allen
24-11-1997	Prof. dr. J. Peschar	RUG	Heijke, Van der Velden
24-11-1997	dr. G. van der Werff	GION	Heijke, Van der Velden
5-12-1997	Prof. dr. B. Hövels	ITS	Allen
10-12-1997	drs. J. Groos	OC&W	Allen, Van der Velden
19-12-1997	dr. P. van Oyen	OC&W	Allen
7-1-1998	dr. P. Creemers	CBS	Allen, Van der Velden

Bijlage 2: Geraadpleegde documentatie

Begroting PRIMA 1996/1997. SGW Interne memo nr. 575-08.044/96-3-5C (1996).

Boer, Paul den, Dronkers, Jaap, Esch, Wil van, Hövels, Ben, en Moerkamp, Trudy
Voorstudie potentiële opbrengst individueel onderwijsnummer: Eindrapport. CBS (1996).

De Cohorten VOCL, SGW Interne memo nr. 575-08.044/96-3-5B (1996).

De VOCL-Cohortonderzoeken: Bezinning en Vooruitblik, Verslag van een Studieconferentie VOCL in Den Haag in februari 1996 (1996).

European Longitudinal Studies in the Behavioral and Medical Sciences: Great Britain. Zentralstelle für Psychologische Information und Dokumentation. Informatie op internet (1997).

Hoe verder met prima? Toelichting bij workshop gehouden op 4 december 1997 in Utrecht. SGW Interne memo nr. 575-13.039e/97-2-1 (1997).

Hövels, B.W.M., en Boer, P.A.M. den, *Cohortstudie BVE: Kernpunten van opzet en uitvoering*. ITS/SCO-Kohnstamm (1996).

National Longitudinal Surveys. Frequently Asked Questions. Bureau of Labor Statistics, U.S. Department of Labor, informatie op Internet (november 1997).

National Longitudinal Survey of 1972. National Center for Education Statistics, informatie op Internet (oktober 1997).

PRIMA van SVO naar NWO, SGW Interne memo nr. 575-08.044/96-3-5A (1996).

Riemersma, F., *Activiteiten- en analyseplan. Notitie ten behoeve van SVO* (bijgesteld door A.C.C. van den Oever) SGW Interne memo nr. 575-08.044/96-3-5D (1996).

Roede, Woud, *Prima theorieën over de kwaliteit van het onderwijs*, SVO (ongedateerd).

Wetsvoorstel Invoering Onderwijsnummer Ingediend, VGN-Informatie op Internet (9 januari 1998).

Bijlage 3: Commentaar van experts op het preadvies

- Verzoek om commentaar door Prof. dr. J.A.M. Heijke
- Reactie Prof. dr. J.L. Peschar
- Reactie Prof. dr. J. Hartog
- Reactie Drs. Th. H. Roes
- Reactie Prof. dr. J. Dronkers
- Reactie Dr. C.A. van Bochove
- Reactie Prof. dr. H.B.G. Ganzeboom



Researchcentrum voor Onderwijs
en Arbeidsmarkt

Research Centre for Education
and the Labour Market

Sociaal en Cultureel Planbureau
Prof. dr. P. Schnabel
Postbus 16164
NL-2500 BD Den Haag

Het ROA is ondergebracht bij de Stichting LIBER
van de Faculteit der Economische Wetenschappen
en Bedrijfskunde van de Universiteit Maastricht

ROA is incorporated in LIBER, the Research
Institute of the Faculty of Economics and Business
Administration of Maastricht University

Visiting address
Tongersestraat 53
6211 LM Maastricht

Postal address
Postbus 616
6200 MD Maastricht
The Netherlands

Telephone +31 43 3883 647
Telefax +31 43 3210 999
E-mail secretary@roa.unimaas.nl

Maastricht

Onze referentie
Our reference

Uw referentie
Your reference

Onderwerp
Subject

18 november 1998

ROA/HH/IS/98.527

Geachte heer Schnabel,

Graag nodigen wij u uit commentaar te geven op bijgaand preadvies dat wij hebben samengesteld op verzoek van de Programmaraad Onderwijsonderzoek (PROO).

De PROO is een bij NWO ondergebrachte commissie die zorgdraagt voor een omvangrijk programma van wetenschappelijk onderzoek op het terrein van het onderwijs. Dit programma werd tot enkele jaren terug verzorgd door de Stichting voor Onderwijsonderzoek (SVO).

Een aanzienlijk deel van het programma betreft de zogenaamde cohortenstudies, zoals die in het basisonderwijs (PRIMA) en in het voortgezet onderwijs (VOCL). De PROO beraadt zich op de situatie op de lange termijn met betrekking tot deze cohorten en heeft ons gevraagd hierover een preadvies uit te brengen.

In dit preadvies nemen wij een aantal ontwikkelingen in beschouwing die op lange termijn van grote betekenis kunnen zijn voor de opzet en mogelijkheden van het cohorten-onderzoek. In het bijzonder hebben wij gekeken naar het bredere kader waarbinnen analyses over de effecten van het gevolgde onderwijs geplaatst kunnen worden. Voorts hebben wij ons afgevraagd welke mogelijkheden worden geboden door de voortgaande automatisering van data- en registratiebestanden in samenhang met het in te voeren onderwijsnummer.

Wij zijn ons ervan bewust dat ons advies een bondig karakter draagt en nog vele vragen onbeantwoord laat. Het leek ons echter nuttig eerst af te tasten of het verder vervolgen van de geschetste weg zinvol is.

Hoewel u geheel vrij bent in de keuze van de aspecten waarop u commentaar wil geven, is het van belang zowel de punten te vernemen waar u zich niet in kunt vinden als hetgeen u kunt onderschrijven. Ook uw adviezen zijn uiteraard welkom. Verder geven wij u in overweging in uw commentaar aandacht te schenken aan de privacyproblematiek verbonden aan het gebruik van aan het onderwijsnummer gekoppelde onderwijsprestatiegegevens voor wetenschappelijk onderzoek, de haalbaarheid van de voorgestelde data infrastructuur en de mogelijke wetenschappelijke betekenis ervan.

Dit commentaar is eveneens gevraagd aan prof. dr. J. Dronkers (UvA), prof. dr. J. Hartog (UvA), prof. dr. J.L. Peschar (RUG), prof. dr. H.B.G. Ganzeboom (UU) en dr. C.A. van Bochove (CBS).



Uw commentaar zal worden opgenomen in de definitieve versie van ons preadvies. Graag ontvangen wij uw reactie voor 1 december 1998. Wij zullen over enkele dagen met u contact opnemen om te vernemen of u op ons verzoek om commentaar te geven wil ingaan. Uiteraard hopen wij, gezien uw grote deskundigheid op dit gebied, dat dit het geval is. Uw commentaar is voor de PROO van groot belang om zijn standpunt ten aanzien van de lange-termijnsituatie van de cohortstudies te kunnen bepalen.

Met vriendelijke groet,

Researchcentrum voor
Onderwijs en Arbeidsmarkt

Prof.dr. J.A.M. Heijke
Directeur

Grote Rozenstraat 31
9712 TG Groningen
tel.: 050-3636283

De heer Prof.dr. J.A.M. Heijke
ROA
Postbus 616
6200 MD Maastricht

Datum: 30 november 1998
Ons Kenmerk: sl/jp/144/98

Geachte heer Heijke, *beste Hans*

In antwoord op uw verzoek om commentaar op de concept-nota 'Naar een nieuwe fase in het longitudinale onderwijsonderzoek' bericht ik u graag het volgende.

De hoofdlijn van het stuk spreekt me erg aan. Het is een goed opgebouwd argument en de conclusies stroken met een eerder verhaal dat Anton Wesselingh en ik op de Onderwijssociologische conferentie in 1997 hebben gepresenteerd. (zie bijlage)



Het wezenlijke vraagstuk ligt in de koppelingsproblematiek. In de eerste plaats is er zoveel verschillende informatie dat de collectie inefficiënt wordt. In dit opzicht alleen al is er een noodzaak tot onderlinge coördinatie. In de tweede plaats is er geen andere manier om over een relevante periode longitudinale gegevens te verkrijgen, dan door verschillende databestanden te koppelen. In de derde plaats speelt er ook een kostenaspect: er is veel ondoelmatige overlap in de data en de tijd die het onderzoekers en scholen kost om de data bij elkaar te krijgen.

Natuurlijk is ook het privacy aspect van belang. Terecht wordt daarop gewezen en ook op de maatregelen die in Denemarken in dit opzicht zijn gerealiseerd. Wellicht is het dienstig hier nog even erop te wijzen dat in Zweden sinds het midden van de jaren 70 reeds de Datainspektionen bestaat: die regelt alle aanvragen voor het gebruik van databestanden en de daarop gewenste koppelingen. Op basis van die ervaringen konden Torsten Husen en Ingemar Fågerlind hun fameuze Malmö-studie ontwikkelen: een bestand van 1544 kinderen die op 10-jarige leeftijd in 1938 waren getest en waaraan later op vele manieren gegevens zijn toegevoegd. Zodoende kon een groot deel van de levensloop worden geconstrueerd en is onder meer het effect van adult education op de levensloop onderzocht. (bijgaand de titels van de dissertaties van Fågerlind en Tuijnman)

Tenslotte is het wellicht relevant om ook eens na te denken over de vraag wie er staks de beheerder van een dergelijke datastrategie zou moeten gaan worden (NWO, ministeries, CBS, KNAW?) en wie er toegang toe krijgen. Met name op dit laatste punt zou ik voorstander zijn van een zeer ruim gebruik, net zoals dat in de Amerikaanse longitudinale studies is geregeld. Met name van belang is dat onderzoekers een grote mate van vrijheid krijgen om relevante vragen aan te pakken en dat er ook contra-expertises kunnen worden uitgevoerd door "lastige" onderzoekers.

In de hoop hiermee op uw verzoek voldoende te zijn ingegaan, met vriendelijke groet


Prof. dr. J.L. Peschar



FACULTEIT DER ECONOMISCHE WETENSCHAPPEN EN ECONOMETRIE
Algemene Economie

Prof. dr. J.A.M. Heijke
Researchcentrum voor Onderwijs en Arbeids-
markt
Postbus 616
6200 MD Maastricht

*Roetersstraat 11
1018 WB Amsterdam*

30 november 1998

Afdeling:	Algemene economie
Contactpersoon:	Prof.dr. J. Hartog
Telefoon:	020 - 525 4298
Fax:	020 - 525 4254

Beste Hans,

Hierbij mijn reactie op het preadvies dat je hebt samengesteld op verzoek van de Program-
maraad Onderwijsonderzoek (PROO). Ik vind de basisgedachte heel zinnig maar ik heb wel
een aantal kanttekeningen.

1. Ik zou naar de kosten kijken, in relatie tot de omvang van het databestand. Als
volledige registratie (met alle voordelen van dien) duur is, kun je aan een grote
steekproef denken. Wellicht heb je aan een beperkte jaarlijkse cohortomvang ook
genoeg (10.000 leerlingen?), wellicht zijn de marginale kosten zo laag dat je net zo
goed volledig kunt registreren.
2. Je spreekt van een plan voor de middellange termijn, maar met betrekking tot de
arbeidsmarkt positie en de relatie tot onderwijs is dit eigenlijk een lange termijnplan:
het duurt wel even voor je iets te weten komt over "onderhoud en slijtage" en gebruik
van menselijk kapitaal.
3. Als je iets wilt opbouwen en toch snel beginnen, kun je natuurlijk al cohorten op de
arbeidsmarkt volgen alsof je hun onderwijsgegevens had (je vraagt gewoon terug), je
bouwt dan een systeem waarmee je steeds betere gegevens binnenschuift, en je kunt
daarvan waarschijnlijk zelfs de effecten meten (bijv. retrospectieve onderwijsopgave
vergelijken met registratie tijdens de schooljaren).
4. Koppeling aan het SOFI-nummer (of zelfs gelijkstellen) lijkt me voortreffelijk. Ik heb
niet zulke privacy-zorgen, al is het evident dat je een goed systeem van beveiliging
moet maken, en van toegangsregels.
5. Kun je bij de opbouw teruggrijpen op materiaal dat er lang is? CITO toetsen worden
al tientallen jaren afgenomen, bestaat dat materiaal nog? Daar zou je dan mooi mee
kunnen starten.

Tot zover, hopelijk kun je er iets mee.

Met vriendelijke groet,

Joop Hartog.

UNIVERSITEIT VAN AMSTERDAM

Sociaal en Cultureel Planbureau



Postbus 16164
2500 BD Den Haag

Bezoekadres:

Parnassusplein 5
2511 VX Den Haag

Researchcentrum voor Onderwijs en
Arbeidsmarkt
T.a.v. prof. Dr. J.A.M. Heijke
Postbus 616
6200 MD Maastricht

uw kenmerk ROA/HH/IS/98.527
uw brief van 18 november 1998
ons kenmerk SCP-U.980533.2
onderwerp Pre-advies longitudinaal onderwijsonderzoek

doorkiesnummer (070) 340 70 00
fax (070) 340 70 44
e-mail:

Den Haag, 6 december 1998

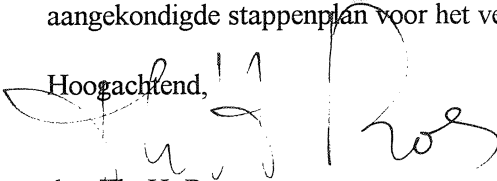
Geachte heer Heijke,

Hierbij wil ik mijn waardering uitspreken voor het preadvies van het ROA over de toekomst van het longitudinale onderwijsonderzoek. In dit preadvies wordt gepleit voor een solide, voor veel doeleinden geschikte data-infrastructuur, waarbij de nadruk niet ligt op individuele onderzoeken, databestanden en beleidsvragen, maar op gebruik van bestaand materiaal, koppeling van bestanden en meervoudige gebruiksdoelen. Dit spoort in belangrijke mate met de wijze waarop het SCP tegen de data-infrastructuur op sociaal en cultureel terrein aankijkt, recente initiatieven bij het CBS op het terrein van een Sociaal Statistisch Bestand en met internationale ontwikkelingen rond Large Scale Facilities op het terrein van de data-infrastructuur. Het door u geschetste perspectief zou voor het SCP een waardevolle bron van informatie zijn. Ik ondersteun het idee dus van harte.

Ondanks bovenstaande instemming wil ik niet verhehlen dat ik mij zorgen maak over de realiseerbaarheid van de plannen. Zo horen bij leerlingnummers adressen die moeten worden gemuteerd. Het muteren van een dergelijk bestand vereist zeer veel inspanning. Koppeling met het GBA kan de zaak vereenvoudigen, maar dan kunnen zich weer bezwaren met betrekking tot de privacy-wetgeving voordoen. Een ander probleem leveren de scores op de CITO-eindtoets basisonderwijs op. Het CITO bewaakt deze gegevens uiterst nauwgezet en het is tot nu toe onmogelijk hiervan voor onderzoeksdoeleinden gebruik te maken.

Twijfels aan de directe uitvoerbaarheid lijken mij echter geen reden om niet te streven naar een data-infrastructuur ten behoeve van longitudinaal onderwijsonderzoek. Organisatorische maatregelen en restricties in de toegang tot databestanden kunnen garanderen dat de privacy niet in het geding komt. Steeds meer wint de overtuiging terrein dat goed onderzoek gebaat is bij zorgvuldig gebruik van bestaande, zo mogelijk gekoppelde gegevens. Steeds meer worden voor problemen rond toegankelijkheid en privacy praktische oplossingen gezocht. Ik kijk dan ook met belangstelling uit naar het aangekondigde stappenplan voor het verwezenlijken van de voorgestelde data-infrastructuur.

Hoogachtend,


drs. Ph. H. Roes
adjunct-directeur

MAX-PLANCK-INSTITUT FÜR BILDUNGSFORSCHUNG

MAX PLANCK INSTITUTE FOR HUMAN DEVELOPMENT

Forschungsbereich Bildung, Arbeit
und gesellschaftliche Entwicklung

Lentzeallee 94
D-14195 Berlin (Dahlem)
Telefon: 030/824 06 - 0
Durchwahl: 030/824 06 -
Telefax: 030/824 99 39

Prof. dr. H. Heijke
ROA
Universiteit Maastricht
Postbus 616
6200 MD Maastricht

15. Dezember 1998

Beste Hans,

Hartelijk dank voor het toesturen van het Preadvies.

Ik zal mijn reactie kort houden in verband met mijn sabbatsverlof. Ik vond echter het preadvies te belangrijk om niet te reageren.

1. De stap naar een data-infrastructuur is goed en logisch. Maar de schets is onvolledig. Een essentieel onderdeel van een dergelijke infrastructuur is ook de terbeschikkingstelling van de daarin aanwezige data en de creatie van analysemogelijkheden. Dat onderdeel is essentieel, want door privacy wetgeving en het normale gedrag van overheden ten aanzien van data en onderzoek is een dergelijke terbeschikkingstelling niet vanzelf sprekend, juist niet voor wetenschappelijk onderzoek dat niet aan de hand loopt van de wensen van de politiek en beleid. Kortom ook het onderzoek moet een duidelijke positie hebben in een dergelijke structuur: dat mag niet overgelaten worden aan het CBS of het SCP om dat te organiseren. Daar zijn die instellingen niet op ingericht. In feite ligt hier een taak voor NWO en/of KNAW. Een dergelijke wetenschappelijk onderdeel van de infrastructuur is ook nodig om te zorgen dat data ook echt gebruikt worden voor wetenschappelijk onderzoek. De oudere longitudinale bestanden zijn alleen veel gebruikt door het bestaan van de toenmalige SISWO werkgroep 'longitudinaal school- en beroepsloopbaanonderzoek'. Een eerste voorwaarde van een dergelijk wetenschappelijk onderdeel van de infrastructuur is een geormeerkte NWO, OCW en SCPB geldstroom voor analyses met en aanvullingen van de aanwezige bestanden.

2. Ik ben somberder over de bruikbaarheid van het onderwijsnummer dat het preadvies op blz. 13. Dat komt niet doordat ik niet de principiële voordelen van dat nummer ziet. Dat komt omdat ik pessimistisch ben over de praktijk die rond dat onderwijsnummer zal ontstaan. De regering zal in het kader van de privacy concessies moeten doen op het punt van de toegankelijkheid van de gegevens van het onderwijsnummer. Het onderzoek, en met name het onderzoek dat niet onmiddellijk nauw verbonden is met een concrete overheidsopdracht

A:\HEIJKE.TXT

zal een veel geringere toegang tot de gegevens van het onderwijsnummer krijgen. De verwijzing naar Denemarken in de voetnoot op blz. 13 is wat dat betreft omineus. Er mag daar dan wel een vergelijkbare infrastructuur bestaan, gelijk heeft men in Denemarken weten te regelen dat de gegevens nooit over scholen met naam en toenaam gepubliceerd mogen worden. Privacy angsten van burgers kunnen heel goed gebruikt worden om ten onrechte belangen van publieke instituties (scholen) te beschermen. Ook de gebeurtenissen rond de Trouw publicatie van inspectiegegevens overscholen illustreert dit duidelijk. Wat ik nu aan mogelijke regelgeving heb gezien rond het onderwijsnummer (zie den Boer, Dronkers, Esch, Hovels en Moerkanp) stemt mij niet optimistisch. Er moet op dit punt een onderscheid gemaakt worden tussen de verkooppraatjes van de betrokken ambtenaar of politicus en de uiteindelijke kleine lettertjes in wet- en regelgeving.

3. De achtergrondgegevens bij het onderwijsnummer zijn volstrekt onvoldoende.

Nationaliteit als indicator van etnische achtergrond is al slecht, het is volstrekt onvoldoende om de verschillen in sociaal-economische en sociaal-culturele achtergrond te meten. Hier moet dus` echt iets op gevonden te worden, wat niet makkelijk zal zijn gezien de zogenaamde privacy gevoeligheden. Ik denk namelijk dat de wetgever om die reden niet bereid zal zijn de lijst van benodigde gegevens uit te breiden voor een voldoende meting van sociaal-economische achtergrond. De realiseerbaarheid van dit punt is een goede test-case voor de haalbaarheid van de richting van het preadvies. Dat OC&W ambtenaren (en ook sommige onderzoekers) op het ogenblik bij achtergrondgegevens alleen denken aan etnische achtergrond spiegelt alleen hun fixatie op het huidige beleid dan de stand van het wetenschappelijk onderzoek (zie bv. van speciale nummer van de TOR waarvan wij de redacteurs waren).

3. Hoewel ik niets tegen de karakterisering 'proces van opbouw, onderhoud en slijtage van menselijk kapitaal' heb (want het past goed bij de tijdgeest en dat brengt geld en steun op), wil ik er toch op wijzen dat er ook andere wetenschappelijk even belangrijke tradities bestaan die een dergelijke data-infrastructuur van een ander perspectief even adequaat kunnen beschrijven: bijvoorbeeld reproductie en verandering van sociale ongelijkheid.

Met vriendelijke groet,

Jaap Dronkers



Centraal Bureau voor de Statistiek

Divisie Sociaal-Economische Statistieken

Statistics Netherlands

cc Rolg
Jim

Researchcentrum voor Onderwijs en Arbeidsmarkt
Prof. dr. J.A.M. Heijke
Directeur
Postbus 616
6200 MD Maastricht

contactpersoon
contact

doorkiesnummer
direct line

uw brief van
your letter of 18 november 1998 ROA/HH/IS/98.527

ons kenmerk
our reference 03265-98-SEC

bijlage(n)
enclosure(s)

onderwerp
re

Voorburg, 29 december 1998

Geachte heer Heijke,

Met belangstelling heb ik kennis genomen van het preadvies aan de PROO over de inrichting van het longitudinale onderzoek op het onderwijsterrein. Kern van dit advies is dat fase-specifieke cohorten worden vervangen door een geïntegreerd en langlopend basispanel dat zich uitstrekt over een langere dan alleen de onderwijsperiode. De informatie in dit basispanel wordt aangevuld met cross-sectioneel onderzoek en met longitudinale deelonderzoeken. Inhoudelijk worden de cohortstudies opgevat als deel van een breed onderzoeksterrein dat meer omvat dan alleen de onderwijsperiode.

Zoals bekend heeft het CBS in het meerjarenprogramma *Van cijferfabriek naar knooppunt op de elektronische snelweg* het voornemen geformuleerd een langlopend panel op te zetten en te onderhouden. In dat kader wordt onder meer gedacht aan de mogelijkheid om leerlingen uit de onderwijscohort-onderzoeken te laten instromen in het Sociaal-economisch Panelonderzoek (SEP). Op langere termijn gaat het om een langlopend panel op basis van registraties aangevuld met enquêtegegevens. De informatievoorziening op het gebied van sociaal-economische dynamiek heeft dan ook een prominentere rol gekregen in het werkprogramma van het CBS. In de toekomst zal de vraag naar dit soort informatie nog verder toenemen, waarbij vergelijkbaarheid van uitkomsten in de tijd van essentieel belang is. Daarin vormen de onderwijscohortonderzoeken een belangrijk element.

Voor het maken van een database waaruit informatie kan worden samengesteld over de levensloop van mensen, zal het CBS op de middellange termijn steeds minder gebruik maken van data die door middel van enquêtering zijn verkregen. Als eenmaal het onderwijsnummer is doorgevoerd, is het mogelijk een grotendeels administratief cohort te onderhouden waardoor de enquêtedruk bij scholen en bij huishoudens aanzienlijk kan worden teruggedrongen. Beschikbaarheid van koppelbare basisgegevens over onderwijs stelt het CBS in staat om zonder administratieve lastendruk en zonder hoge kosten statistische informatie samen te stellen over de sociaal-economische en etnische achtergronden van (de ouders van) leerlingen in relatie tot hun schoolprestaties; de doorstroom van leerlingen naar de arbeidsmarkt, inclusief de bedrijfsklasse, arbeidsduur, hoogte van het loon, aard van het dienstverband en latere loopbaan, alsmede het eventuele beroep op de sociale zekerheid.

CBS Voorburg
Prinses Beatrixlaan 428
Postbus 4000
2270 JM Voorburg

Tel. 31 (0)70 337 38 00
Fax 31 (0)70 387 74 29

CBS Heerlen
Kloosterweg 1
Postbus 4481
6401 CZ Heerlen

Tel. 31 (0)45 570 60 00
Fax 31 (0)45 572 74 40



Vervolgblad nr. 1 behorende bij brief 03265-98-SEC d.d. 29 december 1998

In een later stadium zouden ook over de inkomensontwikkeling en aspecten van de leefsituatie, statistische gegevens kunnen worden opgenomen. De bronnen die het CBS daarbij zal benutten zijn onder meer de bevolkingsadministratie (GBA), de verzekerdenadministraties, de administraties van de uitvoeringsinstellingen sociale zekerheid, de door het CBS uit loon- en salarisadministraties verkregen gegevens, bestanden van de arbeidsbureaus, enquêtebestanden van het CBS zelf zoals de Enquête beroepsbevolking, en het Permanent Onderzoek naar de Leefsituatie. Uiteraard kunnen deelsteekproeven uit dit administratief panel worden benaderd voor aanvullende bevraging over onderwerpen waarover uit registraties geen informatie beschikbaar is.

De ideale opzet van de data-infrastructuur wijkt in onze optiek dus enigszins af van het schema op pagina 11 in de nota. Gegevens uit registraties vervullen een meer prominente rol. Het CBS rekent het tot zijn eigen taak een basisregistratie aan te leggen en die aan te vullen met 'duurzame' kerninformatie, voor elementen die genoemd zijn in het aanvullend onderzoek kan gebruik worden gemaakt van andere bronnen die bij het CBS voorhanden zijn

Het preadvies doet geen uitspraken over de concrete organisatorische opzet van de voorgestelde data-infrastructuur. Gezien het maatschappelijk belang van de beschikbaarheid van de gegevens op de kortst mogelijke termijn, de maatschappelijk functie van het CBS, de beschikbare databronnen en het feit dat de problemen rond het koppelen van databestanden grotendeels zijn opgelost, ligt het voor de hand het CBS een centrale rol toe te kennen bij het opzetten van een onderwijs-database. De Wet op het CBS en de Centrale Commissie voor de Statistiek regelt in artikel 9 dat het CBS gebruik kan maken van alle registraties bij instellingen en diensten van het Rijk, in overeenstemming met de Minister die voor deze administraties verantwoordelijk is, en daarbij bevoegd is het SOFI-nummer voor statistische doeleinden te gebruiken. Het CBS heeft uiteraard technische en organisatorische voorzieningen getroffen om de geheimhouding te waarborgen. Over het geheel van beveiligingsprocedures wordt regulier overlegd met en geadviseerd door de Registratiekamer.

Hoogachtend,

Directeur Divisie
Sociaal-economische Statistieken,

(DR. C.A. VAN BOCHOVE)

Uw kenmerk

Ons kenmerk

Doorkiesnummer

**Opleiding Sociologie
Onderzoekscentrum ICS/U**

Datum

Onderwerp

9 maart 1999

prof. dr. H. Heijke
ROA, Universiteit van Maastricht
Postbus 616
6200 MD Maastricht

Heidelberglaan 1, De Uithof
Postbus 80.140
3508 TC Utrecht
Telefoon (030) 253 2101/4074
Fax (030) 2534405
Email: H.Ganzeboom@accu.uu.nl
[Http://www.fss.uu.nl/soc/hg](http://www.fss.uu.nl/soc/hg)




Faculteit Sociale Wetenschappen

Universiteit Utrecht

Geachte Heer Heijke,

Hierbij doe ik U het per 18 november jl. gevraagde advies toekomen. Het spijt me dat het zo lang op zich heeft laten wachten.

Met vriendelijke groet,


Harry B.G. Ganzeboom

NAAR EEN NIEUWE FASE IN HET LONGITUDINALE ONDERWIJSONDERZOEK. PRE-ADVIES, SAMENGESTELD OP VERZOEK VAN DE PROGRAMMARAAD ONDERWIJSONDERZOEK VAN NWO

1. Het pre-advies geeft een adequaat en leerzaam overzicht van het in gang zijnde longitudinale onderwijsonderzoek, nu en in het verleden in Nederland en een goed uitzicht wat er op dit gebied in een aantal andere landen gaande is. Op dit gebied zijn in Nederlands van verschillende kanten in het verleden aanzienlijke onderzoeksprestaties verricht, die het waard zijn gecontinueerd en uitgebreid te worden. Het pre-advies legt de vinger op tenminste één probleem in de bestaande onderzoeksopzetten: de incompleetheid van gegevens die door verschillende oorzaken (paneluitval, verstrooiing van de steekproef, beperking van de onderzochte levensfasen) aan een cohortvolgende onderzoeksopzet kleeft. De voorgestelde oplossing, het creëren van een data-infrastructuur op basis van het in te voeren onderwijsnummer, kan inderdaad een aanzienlijke kwaliteitsslag voor dit onderzoek betekenen.
2. Het pre-advies kiest wat betreft theoretische oriëntatie een wel erg exclusief economische vertrekpunt met de formulering "opbouw, onderhoud en slijtage van menselijk kapitaal". Men zou uit deze formules niet meer vermoeden dat het onderwijs niet alleen de leerlingen toerust met waardevolle competenties en vaardigheden, maar ook een belangrijke bron is van sociale ongelijkheid en sociale uitsluiting. Vragen over de verdelende functie van onderwijs hebben in de afgelopen jaren een sterke toespitsing gekregen op vragen over uitsluiting en integratie van allochtonen. Het lijkt me dat nieuwe onderzoeksopzetten met dit punt expliciet rekening moeten houden.
3. De notitie is in mijn ogen onduidelijk in haar schets van de mogelijkheden en onmogelijkheden van cohort(panel)onderzoek. Anders dan hier wordt voorgesteld is panelonderzoek in het algemeen niet zomaar geschikt om 'maatschappelijke tendensen' (d.w.z. historisch trends) zichtbaar te maken. Strikt genomen is panelonderzoek aan één cohort zelfs niet geschikt om aan geobserveerde veranderingen een levensloopinterpretatie te geven. Om de vereiste scheiding tussen historische veranderingen (cohort, periode) en levensfase-effecten te maken, moet aan nog een voorwaarde voldaan zijn, namelijk dat personen ondervraagd worden die dezelfde levensfase op een verschillende tijdstip of (wat hetzelfde is) op een tijdstip in verschillende levensfasen verkeren. Sommige van de bekende onderwijscohorten voldoen nu juist niet aan die eis, doordat men met een groep is begonnen die in een of enkele jaren geboren zijn. Uit zulke gegevens kan men pas wat over historische trends leren, wanneer de gegevens worden gelegd naast die van een eerder of later cohort. De tragiek van panelonderzoek is dat men lang moet wachten alvorens zulke gegevens ter beschikking komen en dat ze dan vaak lijden onder onvergelykbaarheid, omdat ze nu eenmaal onder een ander tijdstip gestart zijn (en vaak met een ander onderzoeksteam). Voor succesvol gebruik van cohortgegevens is het heel belangrijk dat men binnen een onderzoeksopzet met cohorten van verschillende levensfase begint, en daarbij retrospectieve informatieverzameling niet schuwt. Alleen op die manier krijgt men binnen een aantrekkelijk termijn gegevens over sociale verandering ter beschikking. Een ideale opzet wat mij betreft zou eruit bestaan binnen een panel met verschillende cohorten tegelijk te starten en regelmatig nieuwe cohorten toe te voegen met het verstrijken van de jaren. Ideal zou men moeten starten in verschillende levensfasen (scholierenpanel, jongvolwassenen panel, etc.) om het beeld van de bevolking compleet te houden. Om zo'n onderzoeksopzet homogeen te houden, moet men ook aan belangrijke organisatorische en bestuurlijke voorwaarden voldoen, waarin in iedere geval een centrale groep onderzoekers waakt over de homogeniteit van het design.

4. Aan de voorgestelde opzet van een data-infrastructuur kleven belangrijke privacy-problemen, vooral omdat voor een deel de (basis)informatie integraal bijeen wordt gebracht. Nu ben ik op zichzelf erg blij dat we in Nederland nu eindelijk de koudwatervrees verloren zijn en het tij gekeerd is. Alom lijkt immers het inzicht door te breken dat het systematisch bijeenbrengen van statistische gegevens over de bevolking ook goede doelen dient. Ook meen ik dat een overheid die het collectief welzijn (in dit verband het onderwijs) bevordert, recht heeft op adequate informatie over de effecten van het beleid. Dit alles neemt niet weg dat het voorstel om een basisregistratie van deze omvang en aard op te zetten niet zonder slag of stoot tot stand zal komen. Ook verwacht ik dat de toegang voor onderzoekers nog wel eens ernstig belemmerd zou kunnen worden, ook al zijn de gegevens ter beschikking, vooral wanneer het CBS een grote rol hierbij gaat spelen (hetgeen te voorzien valt). Het lijkt me dat vanuit NWO als initiatiefnemer garanties moeten worden geschapen om toegang voor wetenschappelijk onderzoek mogelijk te maken. Het spreekt vanzelf dat dit moet gebeuren op basis van een harde regeling van het gebruik van de data en expliciete anonymisering hiervan. Dat dit geen vanzelfsprekende zaak is, blijkt wel uit het feit dat ook momenteel gegevens van de CBS cohorten slechts mondjesmaat en met grote verminking ter beschikking komen van onderzoekers.

Harry B.G. Ganzeboom
Hoogleraar Sociologie, UU

Unigraphic-opdrachtnummer:

Josien

Geachte opdrachtgever,

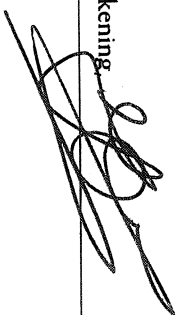
Deze proef is speciaal bedoeld voor
tekstcontrole.

Wij verzoeken U dan ook deze proef goed
op **auteurs- en zetfouten** na te kijken.

Wilt U deze proef getekend voor accoord
retour zenden,

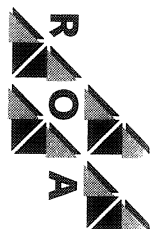
Datum, 31/01/00

Handtekening



Peter Debyeplein 15
6229 HA Maastricht

Telefoon 043 388 3970
Telefax 043 361 3824



Researchcentrum voor Onderwijs
en Arbeidsmarkt
Research Centre for Education
and the Labour Market

Jozien Hendriks
Hoofd secretariaat

Tongersestraat 53
Postbus 616
6200 MD Maastricht
Nederland

Telefoon: +31 43 3883 647
Telefax: +31 43 3884 914
E-mail: j.hendriks@roa.unimaas.nl
Internet: <http://www.unimaas.nl/~roa>

Universiteit Maastricht

oek op
waaktaken
sup.

**Naar een nieuwe fase
in het longitudinale
onderwijsonderzoek**

het longitudinale onderwijsonderzoek

Researchcentrum voor
Onderwijs en Arbeidsmarkt
Postbus 616
6200 MD Maastricht

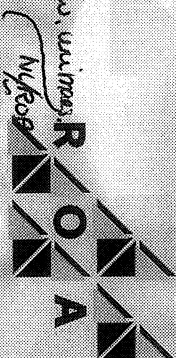
Telefoon: 043 - 3883647
Telefax: 043 - 3884914
E-mail: secretary@roa.unimaas.nl

Universiteit Maastricht
Faculteit der Economische Wetenschappen en Bedrijfskunde

→ nog toevoegen

Internet: <http://www.unimaas.nl>

ROA-R-1999



Researchcentrum voor Onderwijs
en Arbeidsmarkt
Research Centre for Education
and the Labour Market

Internet

uitgel
aanbouders sop

Reflex blue

Formaat: 740x165 mm

Vernieuwing ontwerp en druk Unigraphic